



CIMAT



# Predicción de la producción de energía de una planta de poder de ciclo combinado usando herramientas de AutoML

Ridge and Lasso Regression, AutoKeras y PyCaret.

Tópicos Selectos de Cómputo: AutoML  
Centro de Investigación en Matemáticas, A.C.

Gabriel Alejandro Aguilar Farrera

# Contenido

1. Descripción de la base de datos
2. Objetivo del Proyecto
3. Diseño experimental
4. Conclusiones

# Descripción de la base de datos: Combined Cycle Power Plant

El conjunto de datos contiene **9568** puntos de datos recopilados de una planta de energía de ciclo combinado durante **6 años** (2006-2011), cuando la planta de energía estaba configurada para funcionar a carga completa.

Variables:

- **Temperatura (T)** en el rango de 1.81°C a 37.11°C.
- **Presión Ambiental (AP)** en el rango de 992.89- 1033.30 milibares.
- **Humedad Relativa (RH)** en el rango de 25.56% a 100.16%.
- **Vacío de Escape (V)** en el rango de 25.36-81.56 cm Hg.
- **Salida de energía eléctrica neta por hora (EP)** de 420.26-495.76 MW.

# Descripción de la base de datos: Combined Cycle Power Plant

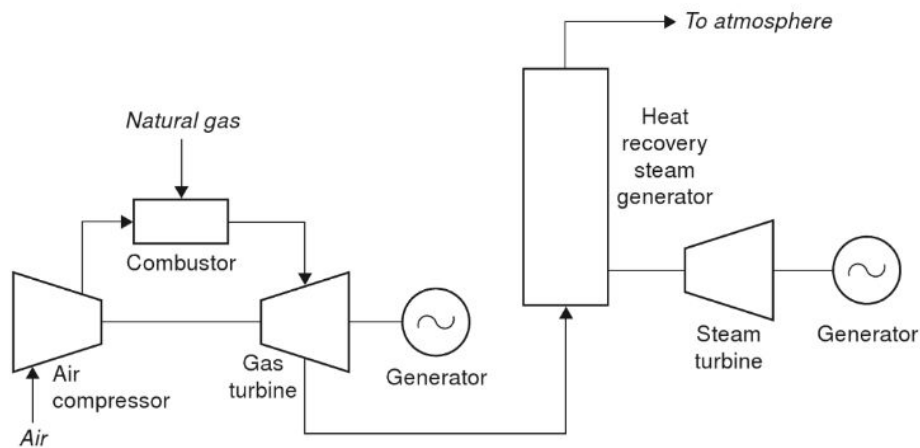


Fig. (1): CCPP.

# Objetivo del Proyecto

1. Identificar el mejor modelo de regresión para este problema.
2. Comparar los resultados obtenidos con los modelos de regresión con los obtenidos con ayuda de AutoKeras y PyCaret.

# Diseño experimental: Modelo de Regresión

- Análisis de correlación lineal
- División de datos: 80% train y 20% test
- Estandarización de los datos
- Rejilla de valores posibles para  $\lambda$
- Selección de  $\lambda$
- Evaluación de los modelos

# Modelo de Regresión: Análisis de correlación lineal

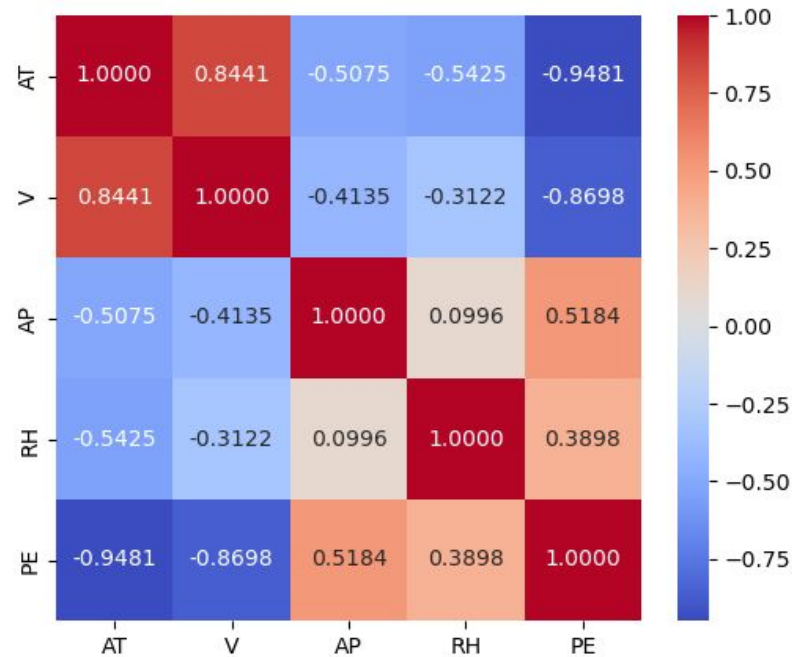


Fig. (2): Mapa de correlación

# Modelo de Regresión: Valores posibles para $\lambda$

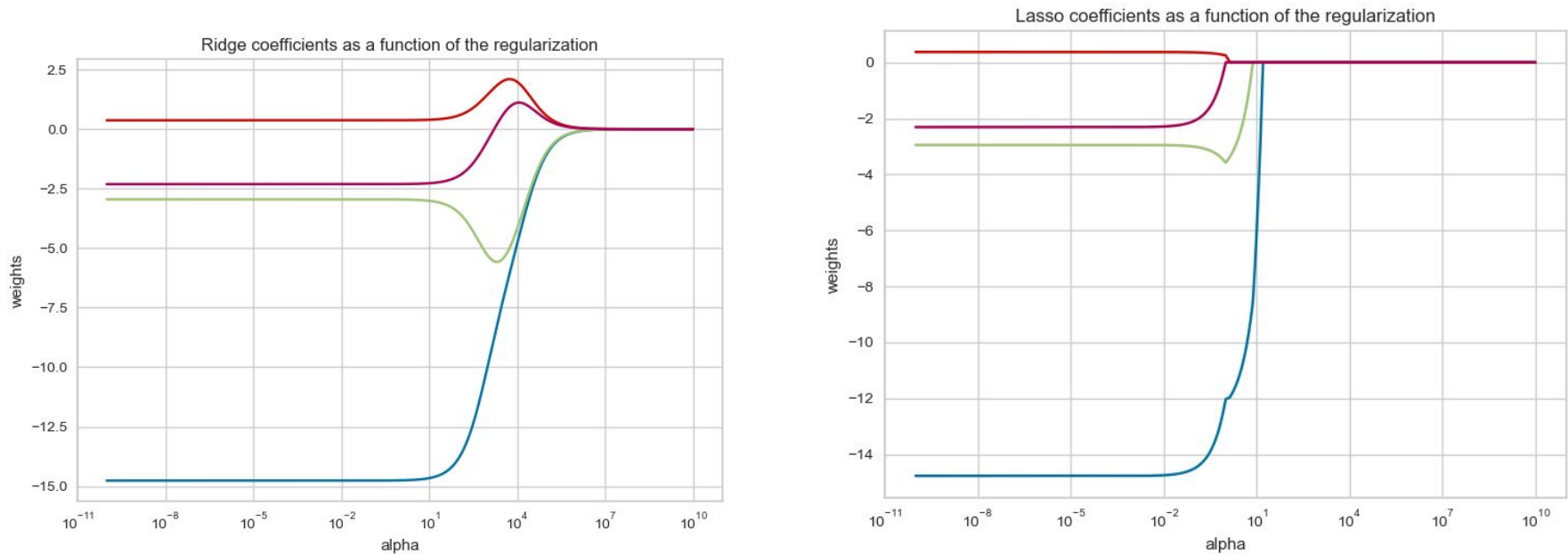


Fig. (3): Coeficientes Ridge y Lasso en función de su valor de  $\lambda$



# Modelo de Regresión: Valores posibles para $\lambda$

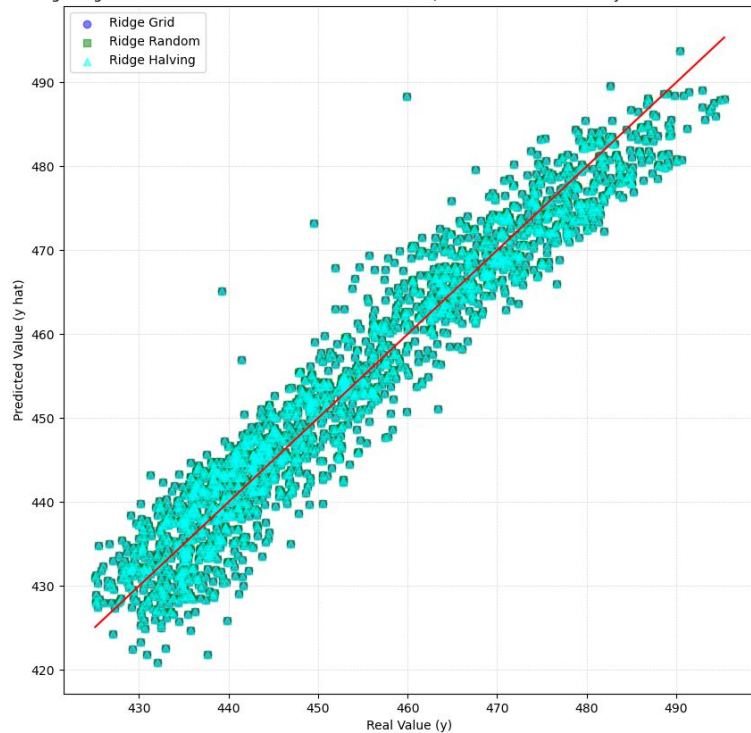
	Regresión Ridge	Regresión Lasso
GridSearchCV	$\lambda = 0.758$	$\lambda = 0.01$
RandomizedSearchCV	$\lambda = 0.735$	$\lambda = 0.010$
HalvingSearchCV	$\lambda = 1.576$	$\lambda = 0.080$

TABLE II: **Datos estandarizados:** Valores de  $\lambda$  encontrados en cada método.

Fig. (4): Coeficientes Ridge y Lasso en función de su valor de  $\lambda$

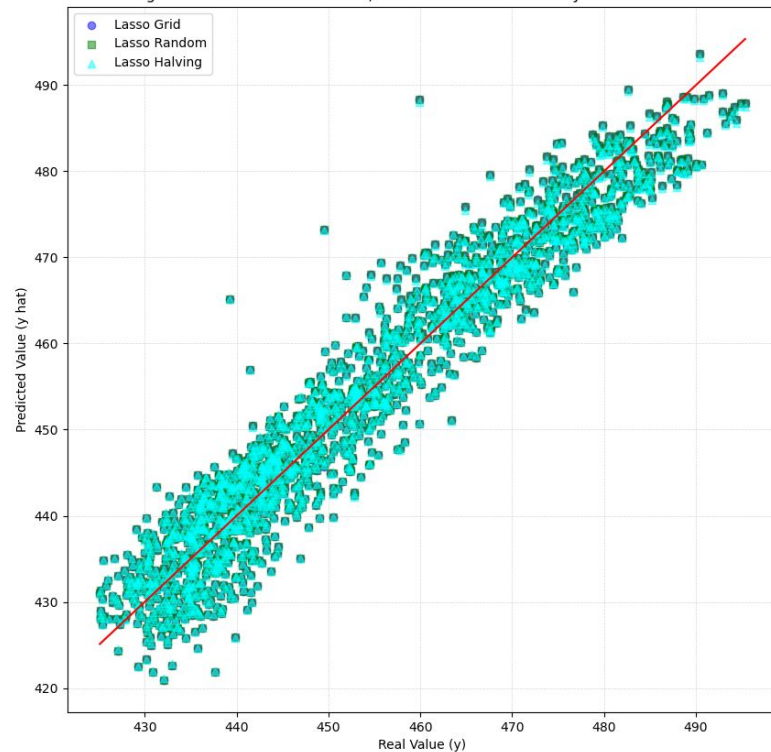
# Modelo de Regresión

Ridge Regression Model with  $\lambda = 0.758047842635104, 0.7353505400985709$  y  $1.5767020233251103$



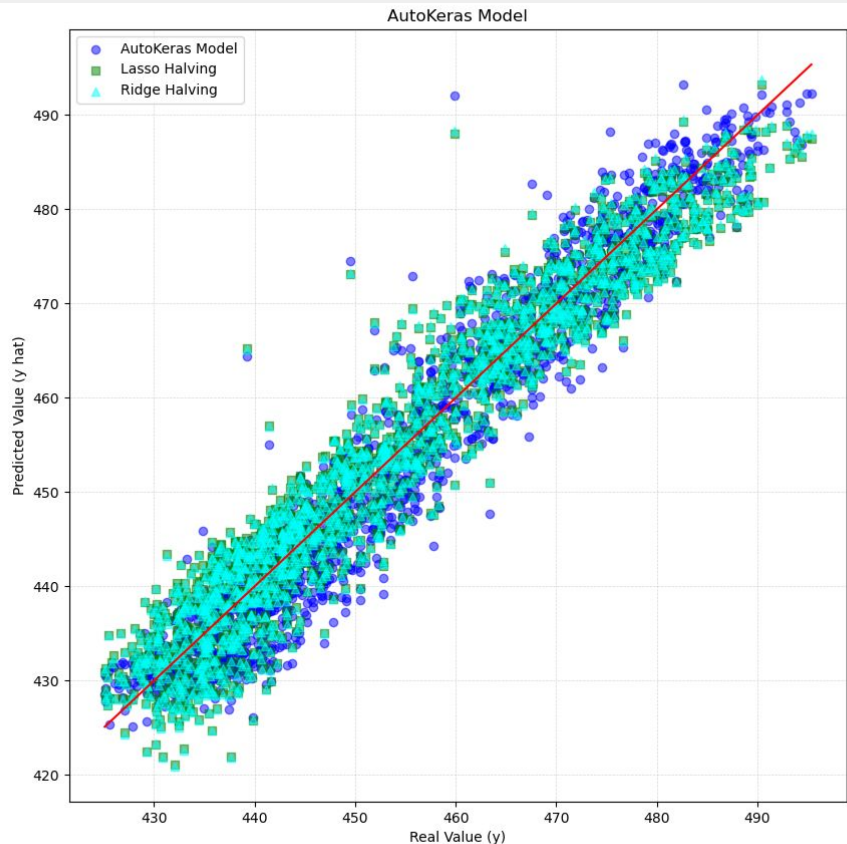
R. Ridge:  $R^2$  test = 0.9301 y MSE test = 20.2719.

Lasso Regression Model with  $\lambda = 0.01, 0.010153156962070463$  y  $0.08002958074821419$

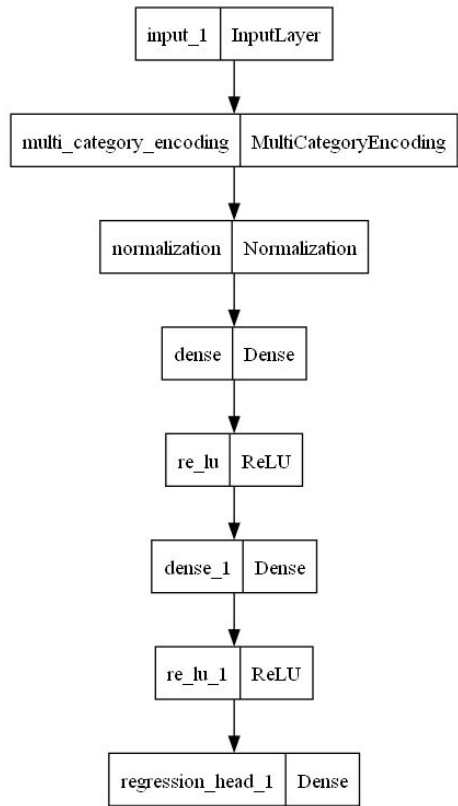


R. Lasso:  $R^2$  test = 0.9301 y MSE test = 20.2663.

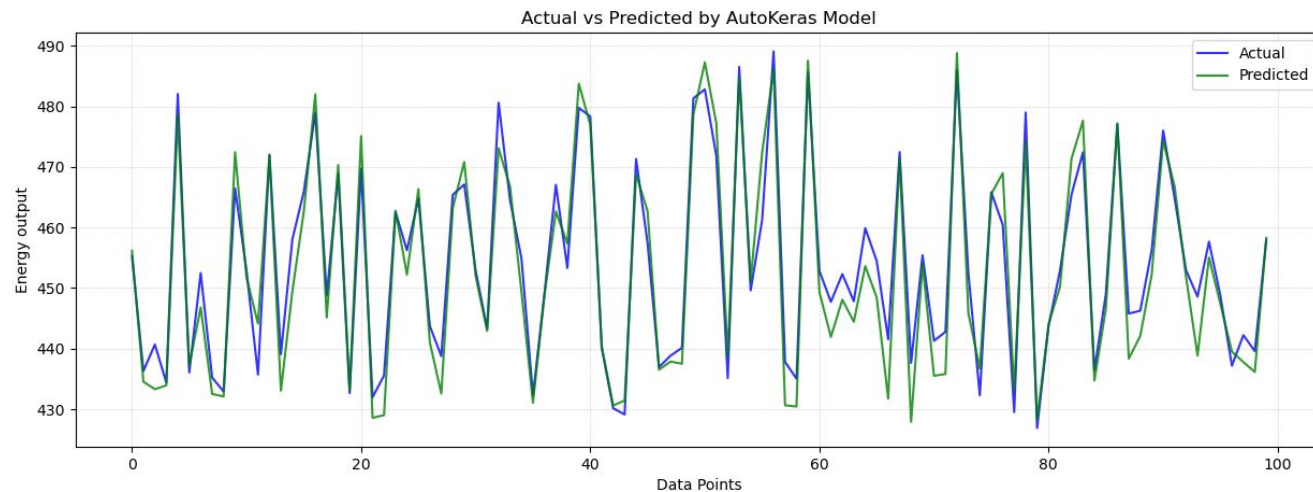
# Modelo de Regresión: AutoKeras



AutoKeras Model: R2 test = 0.9387 y MSE test = 17.7606.



# Modelo de Regresión: AutoKeras



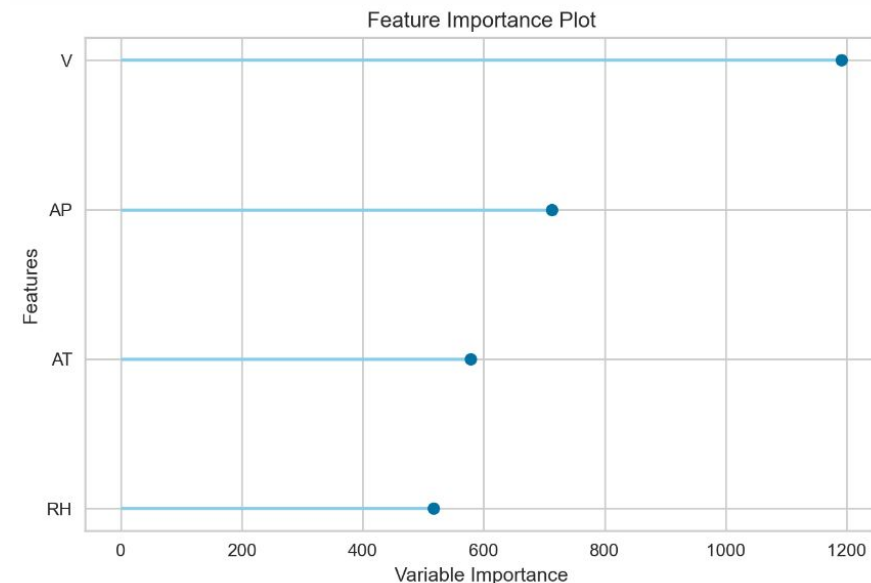
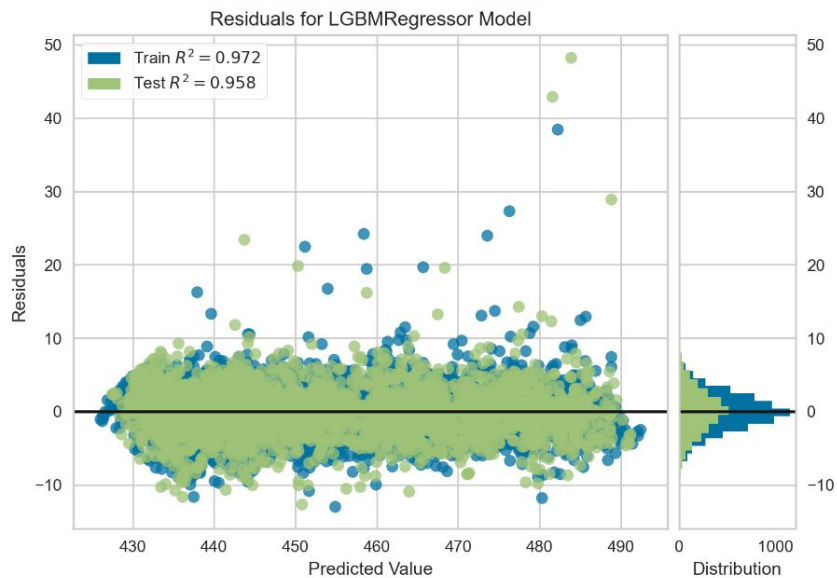
AutoKeras Model:  $R^2$  test = 0.9387 y MSE test = 17.7606.

# Modelo de Regresión:PyCaret

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
<b>xgboost</b>	Extreme Gradient Boosting	2.3694	11.3186	3.3562	0.9609	0.0074	0.0052
<b>lightgbm</b>	Light Gradient Boosting Machine	2.5526	12.1228	3.4759	0.9581	0.0076	0.0056
<b>rf</b>	Random Forest Regressor	2.5241	12.4675	3.5255	0.9569	0.0077	0.0056
<b>et</b>	Extra Trees Regressor	2.5108	12.6813	3.5517	0.9562	0.0078	0.0055
<b>gbr</b>	Gradient Boosting Regressor	3.0171	15.7059	3.9565	0.9457	0.0087	0.0066
<b>knn</b>	K Neighbors Regressor	2.9141	15.9994	3.9960	0.9446	0.0088	0.0064
<b>lar</b>	Least Angle Regression	3.6582	21.2294	4.6030	0.9266	0.0101	0.0081
<b>br</b>	Bayesian Ridge	3.6582	21.2294	4.6030	0.9266	0.0101	0.0081
<b>ridge</b>	Ridge Regression	3.6584	21.2294	4.6030	0.9266	0.0101	0.0081
<b>lr</b>	Linear Regression	3.6582	21.2294	4.6030	0.9266	0.0101	0.0081

Mejores 10 modelos.

# Modelo de Regresión: PyCaret



Los datos se separaron en 6697 para entrenamiento y 2871 para prueba.

# Conclusiones

- Los modelos Lasso y Ridge tienen buenos valores en sus métricas de evaluación ( $R^2$  superior a 0.90).
- La solución óptima de Ridge y Lasso es aquella que no hace contracción de variables.
- AutoKeras es una herramienta de AutoML enfocada en DL y además es fácil de usar.
- Entre los modelos de regresión Ridge y Lasso y las arquitecturas dadas por AutoKeras, las arquitecturas de AutoKeras fueron superiores, sin embargo, a cambio de un modelo mucho más complejo y una mejora despreciable.
- PyCaret es una herramienta sumamente fácil de utilizar y ésta fue la que dio el modelo con los mejores resultados: Extreme Gradient Boosting.

# Referencias

- [1] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An introduction to statistical learning: With applications in python. (No Title)
- [2] Machine Learning Repository: Combined Cycle Power Plan